DOCUMENT RESUME

ED 251 510 TM 850 012

AUTHOR Zwick, William R.; Velicer, Wayne F.
TITLE A Comparison of Five Rules for Determining the Number of Components in Data Sets.
PUB DATE Aug 84
NOTE 59p.; Paper presented at the Annual Meeting of the American Psychological Association (92nd, Toronto, Ontario, August 24-28, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Behavioral Science Research; *Comparative Analysis; *Correlation; *Data Interpretation; Factor Analysis; *Factor Structure; Hypothesis Testing; Research Methodology; Sampling; *Statistical Analysis
IDENTIFIERS Cattell (Raymond B); Chi Square Test (Bartlett); *Eigenvalue Greater Than Unity Rule (Kaiser); Minimum Average Partial Rule (Velicer); Parallel Analysis (Horn); Scree Test

ABSTRACT

A common problem in the behavioral sciences is to determine if a set of observed variables can be more parsimoniously represented by a smaller set of derived variables. To address this problem, the performance of five methods for determining the number of components to retain (Horn's parallel analysis, Velicer's Minimum Average Partial (MAP), Cattell's SCREE, Bartlett's Chi-Square test, and Kaiser's eigenvalue greater than unity rule) was investigated across seven systematically varied factors (sample size, number of variables, number of components, component saturation, equal or unequal numbers of variables per component, and the presence or absence of unique and complex variables). Five sample correlation matrices were generated at each of two levels of sample size from the 48 known population correlation matrices representing six levels of component pattern complexity. The performances of the parallel analysis and the MAP methods were generally the best across all situations. The SCREE test was generally accurate but variable. Bartlett's test was less accurate and more variable than the SCREE test. Kaiser's method tended to severely overestimate the number of components. Recommendations concerning the conditions under which each of the methods are accurate are discussed, along with the most effective and useful methods combinations. (Author/BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED251510

A Comparison of Five Rules for Determining the Number of Components in Data Sets

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

W.R. Zwick

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

William R. Zwick

Rhode Island Group Health Association

Wayne F. Velicer

University of Rhode Island

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Running Head: Comparison of Five Rules

TM 850 012

Paper Presented to the American Psychological Association Convention, 1984, Toronto, Canada

ABSTRACT

The performance of five methods for determining the number of components to retain (Horn's parallel analysis, Velicer's MAP, Cattell's SCREE, Bartlett's Chi-Square test and Kaiser's eigenvalue greater than unity rule) was investigated across seven systematically varied factors (sample size, number of variables, number of components, component saturation, equal or unequal numbers of variables per component, and the presence or absence of unique and complex variables). Five sample correlation matrices were generated at each of two levels of sample size from the 48 known population correlation matrices representing six levels of component pattern complexity. The performance of the parallel analysis and the MAP methods was generally the best across all situations. The SCREE test was generally accurate but variable. Bartlett's test was less accurate and more variable than the SCREE test. Kaiser's method tended to severely overestimate the number of components. Recommendations concerning the conditions under which each of the methods are accurate are discussed, along with the most effective and useful methods combinations.

A Comparison of Five Rules for Determining the
Number of Components in Complex Data Sets

A common problem in the behavioral sciences is to determine if a set of p observed variables can be more parsimoniously represented by a smaller set of m derived variables. Component analysis and factor analysis represent two broad classes of procedures designed to solve this problem. The purpose of the analysis may involve the calculation of m scores to replace the original p observations for each subject or the interpretation of the pxm pattern matrix, $\underset{\sim}{A}$, which provides information for the understanding and interpretation of the original variables. The researchers employing parsimony procedures must make a number of decisions, including choice of method, choice of rotation, and choice of method of calculating the scores. One of the most critical decisions is determining how many factors or components to retain (m). The present paper presents the results of a Monte Carlo evaluation of five alternative methods that have been proposed for determining the value of m.

The determination of the number of components or factors to retain is likely to be the most important decision a researcher will make. Decisions involving choice of method, type of rotation, and type of score will have relatively less impact because of the demonstrated robustness of results across different alternatives in these areas. However, under- or over-extraction will distort the results. The obvious problem of underextraction involves the loss of important information by ignoring a

factor or combining it with another factor. The effects of overextraction, followed by rotation, is less well doctumented but equally important. Comrey (1978) describes some of the dangers, such as minor factors being built up at the expense of major factors and/or the creation of factors with only one high loading and a few low loadings. These are factors that are both unlikely to replicate and will be uninterpretable. Velicer and Jackson (1984) assert that overextraction is likely to be the prime reason for discrepancies between factor analysis and component analysis. In view of how important this decision is to the analysis, it is interesting that some recent textbooks provide little or no guidance in this area (Chatfield & Collins, 1980; Jackson, 1983; and Lunneborg & Abbott, 1983).

Principal Component Analysis (PCA; Hotelling, 1933) may be viewed as involving an eigen decomposition of the P x P sample correlation matrix $\underset{\sim}{R}$, where

[1] $$\underset{\sim}{R} = \underset{\sim}{L}'\underset{\sim}{D}^2\underset{\sim}{L}$$

$\underset{\sim}{D}^2$ is the P x P diagonal matrix containing the eigen roots of $\underset{\sim}{R}$ and $\underset{\sim}{L}$ is a P x P matrix which contains the corresponding eigen vectors. When component analysis is employed as a parsimony model, only the first m components are retained. The component pattern ($\underset{\sim}{A}$) may be written as

[2] $$\underset{\sim}{A} = \underset{\sim}{L}_m \underset{\sim}{D}_m$$

where $\underset{\sim}{D}_m$ contains the first m eigen roots and $\underset{\sim}{L}_m$ contains the corresponding first m eigen vectors. Glass and Taylor (1966), Pruzek and Rabinowitz (1981), and Kaiser (1970) have reported on the widespread use of PCA in this manner. Velicer (1974, 1976a, 1977) and Velicer, Peacock and Jackson (1982) have shown this use of PCA and Factor Analysis result in essentially equivalent solutions.

A second class of procedures, called common factor analysis (CFA) has also been employed to express a set of P variables more parsimoniously as a smaller set. The factor analytic model specifies that a P x P correlation (or covariance) matrix may be accounted for by m common and P unique factors. This model may be expressed as

[3] $$\underset{\sim}{R} = \underset{\sim}{A}\,\underset{\sim}{A}' + \underset{\sim}{U}^2$$

where $\underset{\sim}{A}$ is a P x m pattern matrix and $\underset{\sim}{U}^2$ is the P x P diagonal matrix of weights for the unique factors. It is important to note that m is frequently assumed to be known for the derivation of these factor analysis procedures. Sometimes the maximum likelihood test is employed to test if the assumed number of factors is correct.

Since both CFA and PCA are employed as data reduction techniques, it is important to note some differences between them. The CFA approach requires that m, the dimension of the reduced set of variables, be known prior to the analysis. The value of m may be determined in one of two general ways. In one approach, m may be determined by applying some mehtod to a PCA solution and the result then used in the factor analysis solution. A second approach uses a maximum likelihood test to evaluate the fit for different values of m. Unfortunately, many of the methods applied to the PCA solution provide different results from each other and from the maximum likelihood approach. Further, Jackson and Chan (1980) have discussed numerous computational difficulties with the maximum likelihood approach itself. In addition, an indeterminacy has been identified in the simultaneous estimation of A and U^2 (Guttman, 1954; Schonemann & Wang, 1972; Steiger & Schonemann, 1979). This indeterminacy is inherent in the factor analysis model. In light of difficulties associated with the requirement that m be known a priori, the indeterminacy of the factor model, the computation problems with factor

analysis, the widespread use of PCA, and the general comparability of results across the two methods, this study chose to focus on the PCA procedure.

Properties of Retained Components

The comparison of methods to determine the number of components to retain requires a description of the qualities desirable in a retained component. A review of the properties of principal components, linked with the goal of data summarization, provides such a discription.

Number of substantial loadings. Intuitively, a parsimony application of PCA requires each retained component to contain at least two substantial loadings. Summarizing power is lost when only one variable is represented. Algebraic (Anderson & Rubin, 1956) and statistical (Lawley, 1940, Morrison, 1976) examinations of CFA agree that at least three variables are required before the first factor can be identified. Anderson and Rubin (1956) have further demonstrated that each subsequent identifiable factor must contain at least three non-zero loadings. At a sample level, a minimum of at least three significant loadings are required for factor identification.

Variance accounted for. Principal components analysis proceeds from a correlation matrix, a standardized variance-covariance matrix in which the variance of each original variable is equal to 1.0. The variance of each principal component is equal to the eigenvalue of that component. The sum of all P eigenvalues is equal to P, the number of variables. A 1.0 eigenvalue component, accounts for as much variance as a single variable. Components with eigenvalues near zero provide no summarizing power. A component with an eigenvalue greater than 1.0 provides more summarizing power than an original variable.

Component reliability. Kaiser (1960) and Kaiser and Caffrey (1965)

addressed the issue of component reliability. Noting that a component must be reliable to be useful, Kaiser (1960) argued that the reliability of a component will always be non-negative when the eigenvalue exceeds 1.0. Horn (1969) noted that this approach to reliability includes all P variables regardless of their component loadings. In applied usage, component scores are usually generated as an unweighted sum of those variables with substantial component loadings. Reliability estimates based only on those items contributing to the component score can be quite high even when the component eigenvalue is below 1.0 (Horn, 1969).

The component properties reviewed above can be integrated to present an operational definition of a useful component. Conventional use of PCA as a data reduction technique, combined with algebraic and statistical necessity in CFA dictate that, at the population level, at least three non-zero loadings are required to identify a useful component. At the sample level, three significant and substantial loadings are needed. In order to guarantee non-negative component reliability, retained components are required to have an eigenvalue greater than 1.0. Therefore, all components with three or more substantial loadings and an eigenvalue of greater than 1.0 will be referred to as major components (MJC). Such MJC components would probably be of interest to most investigators. Components which have either (1) less than three substantial loadings but an eigenvalue of 1.0 or greater or (2) components which have more than three substantial loadings but an eigenvalue of less than 1.0 may be of interest to some investigators and will be referred to as minor (MNC) components. Finally, components with both less than three substantial loadings and an eigenvalue less than 1.0 should never be retained and will be referred to as trivial (TC) components. Table 1 summarizes these operational definitions of major, minor and trivial (MJC, MNC, TC) components.

Insert Table 1 About Here

Determining the Number of Components

A number of rules have been suggested to determine the appropriate number of components to retain (Bartlett, 1950, 1951; Cattell, 1966; Crawford, 1975; Everett, 1983; Horn, 1965; Joreskog, 1962; Kaiser, 1960; Revelle & Rocklin, 1979; Veldman, 1974; Velicer, 1976b). These rules often do not give the same results (Anderson, Acito & Lee, 1982; Cattell & Vogelman, 1977; Hakstian, Rogers & Cattell, 1982; Horn, 1965; Linn, 1968; Zwick & Velicer, 1982). Applied researchers are, therefore, often at a loss as to how to proceed. Conflicting research conclusions can be traced to differing methods of defining the correct number of components.

This section will describe the five methods to be evaluated in this study. The methods are: 1) the Bartlett test; 2) the eigenvalue greater than 1.0 rule; 3) the minimum average partial rule; 4) the scree test and 5) the parallel analysis method. There methods were selected for inclusion either because of their widespread use or their extensive theoretical justification.

Bartlett's test (BART). Following Lawley's (1940, 1941) test for maximum likelihood factor analysis, Bartlett (1950, 1951) developed an analagous statistical test for component analysis of the null hypothesis that the remaining P-m eigenvalues are equal. Each eigenvalue is excluded sequentially until the approximate chi-square test of the null hypothesis of equality fails to be rejected. The first m excluded components are retained.

BART appears sensitive to the sample size. Gorsuch (1973) argued that as the number increases, the tests of significance become more powerful and,

therefore, less and less substantial differences between eigenvalues are found to be significant. This can potentially lead to the retention of more components as a function of the number of subjects. In response to this, Horn and Engstrom (1979) have suggested changing the alpha level at different levels of N. However, as the sample size increases, the estimates of population eigenvalues will become increasingly accurate. This increased accuracy leads to smaller observed differences between equal eigenvalues. This increased accuracy may appropriately offset the increased power of the Bartlett test when the population eigenvalues are actually equal. Zwick and Velicer (1982) found the BART test somewhat more accurate with large samples than with small samples.

Eigenvalue greater than 1.0 (K1). Perhaps the most popular, certainly the most commonly employed method, is to retain the components with eigenvalues greater than 1.0. Kaiser (1960) developed the rationale for this method based on arguments focusing on component reliability, pattern meaningfulness and Guttman's 1954) work examining the lower bounds for the number of components in image analysis. Gorsuch (1974) noted that many users follow Kaiser (1960) and employ the K1 rule to determine the number of components rather than as a lower bound as originally presented. Difficulties associated with this use are noted by Mote (1970) and Humpreys (1964) who argued that rotation of a greater number of components resulted in more meaningful solutions. They imply the relatively blind use of the K1 rule may sometimes lead to the retention of too few components.

A number of researchers (Browne, 1968; Cattell & Jaspers, 1967; Horn, 1965; Lee & Comrey, 1979; Linn, 1968; Ravelle & Rocklin, 1979; Yeomans & Golder, 1982; Zwick & Velicer, 1982) however, have found the number of components retained by this method often overestimates the known underlying

component structure. Gorsuch (1974) and Kaiser (1960) report the number of components retained by K1 is commonly between one third and one fifth or one sixth the number of variables included in the correlation matrix. A Monte Carlo study by Zwick and Velicer (1982) supports this result. This relationship of retained components to the number of variables is detrimental to the accurate estimation of the underlying component structure. The K1 method, although commonly used, is believed by some critics to sometimes underestimate and by many others to grossly overestimate the number of components. The latter situation occurs particularly when there are a large (e.g., P greater than 50) number of variables involved.

The Minimum average partial (MAP). Velicer (1976b) has suggested a method based on the matrix of partial correlations. The average of the squared partial correlation is calculated after each of the m components has been partialed out. Where the minimum average squared partial correlation is reached, no further components are extracted and rotated. The average squared partial correlation reaches a minimum when the residual matrix most closely resembles an identity matrix. After that point, the average squared partial will increase. Using this rule, two variables would be expected to have high loadings on each retained component. The method is congruent with the factor analytic concept of "common" factors. Velicer (1976b) points out the method is exact, can be applied with any covariance matrix and is logically related to the concept of factors as representing more than one variable. In a recent study (Zwick & Velicer, 1982), the MAP rule was more accurate in identifying a known number of components than either the K1 or BART rule.

The SCREE test. Cattell (1966) described this rule which is based upon a graph of the eigenvalues. The scree test (SCREE) is simple to apply. The

eigenvalues are plotted, a straight line is fit through the P-m smaller values and those falling above the line are retained. A number of complications may occur including: (1) gradual slope from lower to higher eigenvalues with no obvious break point in the line; (2) more than one break point in the line; and (3) more than one apparently suitable line may be drawn through the low values. Horn and Engstrom (1979) have noted the underlying similarity of the logic of Bartlett's chi square test and the scree method. Both tests are based on an analysis (one statistical, the other visual) of the essential equality of the "remaining" eigenvalues.

The scree test has been most effective when strong components are present with little confounding due to error or unique factors. Tucker, Koopman and Linn (1969) found the scree test to be correct in 12 of 18 cases. Cliff (1970) found it to be accurate, particularly if questionable components are included. Cattell and Jaspers (1967) found the test to be correct in 6 of 8 cases, while Cattell and Vogelmann (1977) reported the test to be accurate over 15 systematically differing analyses. Further, Cliff and Hamburger (1967) found more definite breaks with larger (N = 400 vs. N = 100) sample sizes and Linn (1968) concurred in this conclusion. Zwick and Velicer (1982) found the scree test to be most accurate with larger samples and strong components. They found the scree test to be the most accurate of four methods evaluated across many examples of matrices of known, non-complex, structure.

Use of the scree test always involves issues of interrater reliability. Cattell and Vogelmann (1977) and Zwick and Velicer (1982) have reported good interrater reliability among naive and among expert judges. However, Crawford and Koopman (1979) have reported extremely low interrater reliabilities. The circumstances associated with high and low interrater reliability on the SCREE procedure have not been identified.

Parallel analysis (PA). Parallel Analysis (Horn, 1965), involves a comparison of the obtained, real data eigenvalues with the eigenvalues of a correlation matrix of the same rank and based upon the same number of observations but containing only random uncorrelated variables. This method is an adaptation of the K1 rule. Guttman's (1954) development of upper and lower bounds was based upon population values. Horn (1965) noted that, at the population level, the eigenvalues of a correlation matrix of uncorrelated variables would all be 1.0. When samples are generated based upon such a matrix, however, the initial eigenvalues exceed 1.0 while the final eigenvalues are below 1.0. Horn (1965) suggested that the eigenvalues of a correlation matrix of P random uncorrelated variables should be contrasted with those of the data set in question, based on the same sample size. Components of the matrix of interest, which have eigenvalues greater than those of the comparison random matrix, would be retained. This approach integrates the reliability and data summarizing emphasis of the population based K1 rule without ignoring the effect of sample size.

Horn (1965) presented one example of PA in a PCA problem. He recommended that the comparison eigenvalues be based upon a number of generated random matrices to avoid major sampling errors in the estimation of the eigenvalues. Although there has been no published systematic examination of the PA method with PCA, Richman (personal communication, Oct., 14, 1983) reported a series of simulation studies with the method. PA was found to be very accurate when applied to correlation matricies conforming to the formal factor analytic model. He further reported that PA led to retention of too many components when applied to correlation matricies conforming to the middle model described by Tucker, Koopman, and Linn (1969). The method was more accurate in both cases at larger (N = 500) than at smaller (N = 100) sample sizes.

Humphreys and Montinelli (1975) applied PA to principal axis factor analysis and found the method accurate over a range of examples. Montinelli and Humphreys (1976) developed a regression equation which accurately predicts the eigenvalues of random correlation matrices with squared multiple correlations inserted as the diagonal. Green (1983) utilized this prediction equation to evaluate the performance of factor analysis of binary items. No such prediction equation has been reported for standard correlation matrices.

Variables Affecting Decision Methods

Previously reported research findings on the performance of the decision methods described above indicate that each may be affected by a set of different factors. These factors include sample size, the number of variables, component saturation, component identification and the presence of special types of variables. This study attempted to incorporate each of these influences into a simulation study. Some of the considerations related to each of these factors are presented in this section.

Sample size. Depending upon the decision method employed, it is possible that the number of subjects may affect the accuracy of the decision about the number of components to retain. Sample size is typically determined both by practical applied considerations and the need for accurate estimation. The sample size must be large enough to allow an adequate estimation of the relationships between the variables. On the other hand, in applied settings, large samples may be too expensive to be practical.

Number of variables. With the development of computer technology and software, larger and larger correlation matrices have been submitted to PCA. PCAs of personality inventories at the item level, for instance, often involve 80 to 100 variables. Analysis of 200 variable sets is becoming common. Larger numbers of variables have been reported to increase the accuracy of

some rules while decreasing it for others (Zwick & Velicer, 1982).

Component saturation. Linn (1968) and Zwick and Velicer (1982) have demonstrated that the underlying component saturation affects decision methods. Underlying components made up exclusively of high loadings (e.g., .80) were more likely to be retained, by various decision methods, than components exclusively made up of lower loadings (e.g., .40).

Component identification. The accurate identification of a component may depend upon the number of variables which have non-zero loadings on that component. Components defined by less than three variables are not capable of identification. The impact of unequal numbers of variables per component is unclear for any of the rules under discussion.

Special variables. Complex variables have a nonzero loading on more than one component. Unique variables have only one non-zero loading and no other variable loads substantially on the same component. Component patterns containing both complex and unique variables are believed to occur frequently in applied situations (Tucker et al. 1969). The effect of these types of variables upon the various decision rules is unclear. Complex or unique variables can be expected to lead to the retention of more components by K1 and BART and to make SCREE decisions more difficult. The effect of such variables upon MAP and PA has not yet been examined empirically.

Methods To Be Included

The correct determination of the number of components has been identified as a crucial step in the data reduction application of PCA. There continues to be general disagreement concerning the best method to accomplish this step. This study compares the performance of five decision methods on simulated data sets. These sets incorporate variables expected to influence each method. The K1 method was included because it is so widely used. The

MAP method was included because of its unambiguous solution, its relation to "common factor" concepts and its good performance in a recent study (Zwick & Velicer, 1982). Bartlett's statistical method (BART) was included because it is the only statistical method appropriate for PCA solutions. The scree test (SCREE) was included because of its apparent simplicity and its reported validity. The parallel analysis method (PA) was included because of its unambiguous solution and its reported accuracy. Each of the chosen methods may be differentially affected by the manipulated variables, including sample size, the number of variables, the degree of component identification and saturation, and the presence of unique or complex variables. The robustness of the five rules in question, across these variables, is the central focus of this study.

Method

Method of Data Generation

Studies of the effectiveness of the various decision methods may be categorized into one of two types. Historically, the more common type of study employed real data representing either new work or "classic" data sets. These studies employed some logical criteria concerning the appropriate number of components and compared the performance of the proposed decision method to the logically determined value (e.g., Cattell, 1966; Horn, 1965; Humphreys & Montanelli, 1975; Velicer, 1976). Such studies, in employing an arbitrary logical criterion, may have inaccurately estimated the performance of the decision method in question. More recently, studies of decision rule effectiveness have employed correlation matrices generated from component structures entirely under the control of the investigator (e.g., Anderson, Acito & Lee, 1982; Cattell & Vogelman, 1977; Tucker, Koopman & Linn, 1969; Zwick & Velicer, 1982). These studies have the advantage of a known criterion

against which to measure the performance of the decision methods. They are, however, open to the criticism that the generated matrices, although conforming to a mathematical model, may not represent real data and thus lead to inappropriate conclusions (Tucker, Koopman & Linn, 1969).

The question of a rule's accuracy cannot be examined without a known criterion. Although logical arguments can be mounted to defend the number of components present in some data sets, these arguments are always open to question. For the assessment of the impact of various conditions upon a rule's accuracy, generated data of a known number of components is preferable. The issue of generalization to real data sets is an important but separate issue which may be independently addressed in the particular way the data is generated. This study employed an approach similar to the "middle model" of Tucker, Koopman and Linn (1969).

Procedure

The number of variables (P) to be employed was set at 36 and 72. These values represent small and moderately large data sets and accommodate constraints imposed by the selection of the number of components to be included. Larger sets of variables have been shown to have a positive impact on MAP, BART (Zwick & Velicer, 1982) and SCREE (Cattell & Vogelman, 1982) and a negative impact on K1 (Zwick & Velicer, 1982).

The sample sizes (N) chosen were selected to reflect common, applied usage. They were set as a function of the number of variables. The lower N was set at twice the number of variables. The higher N was set at five times the number of variables. The resulting N's were 72 and 180 in the cases including 36 variables. When 72 variable cases were examined, N's of 144 and 360 were selected. These appear to include a representative range of sample sizes as reported in applied educational and psychological research. Larger

sample sizes have been shown to moderately improve the performance of the MAP, SCREE and K1 methods (Cattell & Vogelman, 1977; Zwick & Velicer, 1982) and to sometimes improve and sometimes weaken the accuracy of the BART method (Gorsuch, 1975; Zwick & Velicer, 1982).

As described above, major components (MJC) are defined as those with three or more substantial loading and an eigenvalue greater than 1.0. Two types of minor components (MNC) are defined. First, those with three or more substantial loadings and an eigenvalue of less than 1.0. Second, those with less than three substantial loading and an eigenvalue greater than or equal to 1.0. Trivial components (TC) are defined as those with less than three substantial loadings and an eigenvalue of less than 1.0.

The number of major components built into the population correlation matrix was 3 and 6 when P was equal to 36, and 6 and 9 when P was equal to 72. These values were chosen to reflect a reasonable range of reported applied usage. They permit a span of an average number of variables per MJC (P/MJC) from 6 to 12. Although this P/MJC is somewhat high, such values are required to permit variability in the number of variables per component, while still meeting the constraint of at least 3 substantial loadings in each MJC.

The distribution of P/MJC was constrained to be either an equal number of variables per MJC or an unequal number. For the cases where an unequal number existed, the number was symmetrically distributed about the mean number of variables per MJC appropriate for that matrix. That is, if P were 36, and m were 6, the average number of variables per MJC would be 6. When P/MJC was planned to be unequal, the number of variables per component was 8, 7, 6, 6, 5, and 4. Similar distributions for other combinations of P and m are presented in Table 2.

Insert Table 2 about here

Component saturation (SAT), the magnitude of the loading of the variables on a MJC, was split between a high of .8 and a low of .5. These values bridge much of the applied range and have been shown (Zwick & Velicer, 1982) to differentially affect four of the decision rules under consideration. Within any one matrix, the component loadings on all major components were equal and either high (.8) or low (.5).

For the purposes of this study, unique variables (UNIQ) were defined as variables which do not load at all on either MJC's or TC's in the population structure. Instead, UNIQ variables represent the only variable loading on one type of MNC. Such an MNC has a population eigenvalue of 1.0. Complex variables are defined here as those variables which load substantially on a MJC but also load minimally (.2) on a second type of MNC in the population structure.

Specific combinations of these variables were constructed. Previous work has indicated that N, P and SAT have an impact upon some of the decision rules. At each level of P and SAT, component patterns were constructed to evaluate the impact of N, P/MJC and a number of combinations of factors. A "Best Case" set was defined for comparison purposes. This first level of complexity (1) had an equal number of variables per major component, no MNC's or TC's present and thus no complex or UNIQ variables. Five other levels of structural complexity were created for comparison to the "Best Case". This was done by (2) including complex variables to create the MNC's, (3) letting the number of variables per MJC become unequal, (4) including unique variables (as many as the number of MJC's) to provide the second type of MNC, (5)

including unique and complex variables together to provide MNC's, and, finally, (6) a level was constructed which included unequal number of variables per major component and both complex and unique variables. The fifth and sixth levels of complexity were included to highlight the possible importance of the very common situation of unequal numbers of variables per MJC.

Data Generation

Population correlation matrices were created for each combination of the 6 (Complexity) x 2 (P) x 2 (SAT) x 2 (m) factors outlined above. Each population correlation matrix was determined as follows:
One appropriate population component pattern (A) was created in accordance with the level of the number of variables factor (P), the level of the saturation factor (SAT), the level of the number of components factor (m) and the level of the complexity factor under consideration. Post-multiplying by its transpose (AA') resulted in a matrix R* (R* = AA'). Substitution of ones in the diagonal of R* introduced error and produced a population correlation matrix R ($R = R^* + D^2$). The introduction of ones in the diagonal of R raised it to full rank, allowing subsequent analysis. Five sample correlation matrices were generated based on each of these population correlation matrices (Montanelli, 1975) at each level of the number of observations (N) factor.

Principal component analysis was then performed on each of the resulting 480 (6 x 2 x 2 x 2 x 2 x 5) sample correlation matrices. At the time this analysis was performed, the number of components to be retained by each of the four calculable rules (K1, MAP, PA and BART) was determined. Horn and Engstrom (1979) have suggested that the alpha level of the BART procedure should be adjusted to compensate for the increased sensitivity of the test with large sample sizes. Three alpha levels were selected for use with the

BART in this study to incorporate Horn and Englstom's (1979) recommendation. The Bartlett tests were therefore performed at alpha levels of .05 (BA), .001 (BB) and .0005 (BC) in all cases. The PA decision was based on the mean eigenvalues of 50 random correlation matrices at each level of P and N.

The SCREE test was performed on computer generated plots of the eigenvalues of each of the 480 matrices. These plots were examined by two raters trained in the SCREE method (Cattell and Vogelman, 1977). The two raters were college graduates who had majored in psychology. Although they were trained in the SCREE procedure they were uninformed of its purpose. The raters were also naive to the exact purpose of the experiment and had no prior applied experience with the SCREE test. The graphs were presented to the raters in different mixed orders. If either rater asked a question about a particular plot, both listened to an explanation based on the instructions given by Cattell and Vogelman (1977). Whenever possible, examples from the Cattell and Vogelman (1977) directions were used to clarify questions. Independently, an experienced expert judge, uninformed as to the details of the experiment but fully familiar with the use of the SCREE test, rated one sample from each of the 96 cells.

The judgment required by the SCREE method raises the question of rater reliability. Table 3 presents the interrater reliability extimates of the mean of the raters' decisions corrected for the number of raters. The reliability estimates are presented at each level of complexity, saturation and the number of variables. The reliability estimates ranged from .61 to 1 00 with a median value of approximately .88.

Insert Table 3 about here

The correlations of the mean of the raters decisions with the expert's judgment ranged from .60 to .90 across the 6 levels of complexity. The median and mean (Fisher Z transformed) correlation of the averaged rater's decision with the expert's judgment were both approximately .80.

Results

Each decision method leads to an estimate of the number of major components (MJC) to retain. The difference between these rule determined estimates of MJC (RMJC) and the known population value of the number of MJC's (PMJC) was the primary dependent variable in this study. This difference was computed as $\underline{d}$ = RMJC - PMJC. The mean difference from the criteria is an estimate of the method's accuracy. Positive $\underline{d}$'s, therefore, indicate overestimations while negative $\underline{d}$'s indicate underestimations. A difference of 0 indicates a correct estimation of MJC. The standard deviation of the difference is an indication of the methods' consistency. Smaller standard deviations indicate more consistent estimates of MJC. The mean and standard deviation of d for each method, under various conditions, are presented below.

The results are alternately presented first for the P = 36 cases and then, in a parallel fashion, for the P = 72 cases. Each level of sample size (Tables 4 and 5), number of variables per major component (Tables 6 and 7) and pattern complexity (Tables 8, 9, 10 and 11) will be summarized within each level of P and SAT. Tables 12 and 13 present the proportion of each method's estimates of MJC which deviated a set amount from the population value. This representation of the distribution of the estimates is also presented at each level of P and SAT.

Table 4 presents the means (d) and standard deviations (sd) of the

difference between each method's estimate of MJC and the known MJC for different sample sizes when P = 36 and the component saturation is .5 and .8. Table 4, therefore, summarizes results collapsed across all six levels of pattern complexity and both levels of the number of variables per MJC in order to allow an examination of the individual impact of sample size. Each of the four rows of differences in Table 4 represent 60 observations. Tables 4 through 11 follow essentially the same format. A detailed description will, therefore, be given only for Table 4.

The first row of Table 4 presents the mean difference of each method's estimate of MJC from the known value when P was 36, the saturation was .5 and the sample size was 72. Under these conditions, the MAP method produced a mean difference score (d) of -1.08, an underestimation. The PA method produced a much smaller underestimation of -0.05. The SCREE (0.50) and K1 (8.32) methods both overestimated the criterion with K1 providing a very large overestimation. The results given by the Bartlett method were calculated for alpha levels of .05, .001 and .0005, as indicated above. At each level, the Bartlett method led to underestimations (-2.87, -3.92, -3.98 respectively). The Bartlett method retained fewer components at the more conservative alpha levels. The standard deviations (sd's) associated with the mean difference scores for each method are presented in row 2. They ranged from 0.70 for the PA method to 1.68 for the MAP method. The third and fourth rows of Table 4 present the mean difference of each method's estimate of MJC from the known value and the standard deviations when P was 36, the saturation was .5 and the sample size was 180. The increase in N from 72 to 180 appeared to have had minor effects on the MAP, PA and SCREE methods. The K1 and BART methods show some improvement at the higher level of N. Rows 5 and 6 of Table 4 present the mean differences and standard

deviations for each method when P was 36, the saturation was .8 and the sample size was 72. All the methods showed improved average estimates of the criterion at this higher level of saturation. It should be noted, however, that the standard deviation of the differences increased for all levels of the BART method and, to a lesser extent, for the K1 rule as well. Rows 7 and 8 of Table 4 present the mean differences and standard deviations for each method when the sample size was increased to 180, P was 36 and the saturation was .8. Compared to the results in rows 5 and 6, the larger sample size resulted in more accurate (d = 0.0) and consistent (sd = 0.0) estimations by MAP and PA methods. The performance of the SCREE and K1 method was not greatly affected. The three levels of the BART method retained more components at the higher sample size. This led to a larger overestimation at BA and a switch from under- to overestimation at BB and BC. The standard deviations at all three levels of BART appear to have been larger at N = 180 than at N = 72.

Insert Table 4 About Here

The K1 method performed slightly better at the higher sample size at both levels of component saturation. BART retained more components at the higher level of sample size at both levels of component saturation. Table 5 parallels Table 4 with P equal to 72. It summarizes the impact of sample size at both levels of component saturation. The MAP and PA methods were again minimally influenced by the sample size change at both levels of component saturation. When the saturation was .5, the SCREE method showed less overestimation at the higher than at the lower sample size. This effect was not apparent when the saturation was .8.

Insert Table 5 About Here

The role of the number of variables is presented from a different perspective in Tables 6 and 7. Table 6 presents a summary of the results for each average number of variables per MJC (P/MJC) when P equaled 36 and the saturation was .5 and .8. At both levels of saturation, MAP, PA and SCREE performed more accurately and consistently when the average number of variables per MJC increased from 6 to 12. K1 showed an increased overestimation when P/MJC increased and the saturation was .5 and a decreased overestimation when P/MJC increased and the saturation was .8. The BART method consistently showed a decrease in the number of components retained as P/MJC increased. When the saturation was .5, this resulted in larger underestimations; while at a saturation of, .8 BART's estimates changed from overestimations to underestimations as P/MJC increased.

Insert Table 6 About Here

Table 7 parallels Table 6 with P equal to 72. The MAP, PA and SCREE methods showed improved performances at the higher level of P/MJC when the saturation was .5. When the saturation was .8, MAP, PA and SCREE showed essentially no improvement at the higher level of P/MJC. The K1 and BART methods showed some improvement at the higher level of P/MJC at both levels of saturation.

Insert Table 7 About Here

Tables 8 and 9 present a summary of the methods' performance when P was equal to 36 and the saturation was .5 or .8 at each of six levels of pattern complexity. The results are collapsed across both levels of N and P/MJC so that each level of Complexity represents 20 observations. As defined above, at Complexity level 1 the MJC's contain equal numbers of variables. There are no unique or complex variables at Complexity level 1 and hence no MNC's or TC's. Level 2 is the same as level 1 except it includes MNC's comprised of low complex loadings. Level 3 is the same as level 1 except the number of variables per MJC is unequal across MJC's. Level 4 is the same as level 1 except it includes MNC's made up of unique variables. Level 5 combines both TC's comprised of complex loadings and MNC's comprised of unique variables. Level 6 is the same as level 5 except the number of varibles loading on each MJC is unequal across major components.

Insert Table 8 and 9 About Here

The range of pattern complexity affected the methods differently. Although the methods tended to perform best at Complexity level 1, they had different worst cases. When the saturation was .5, in Table 8, the worst cases were: MAP and PA at level 5; SCREE at level 2; K1 and BART at level 4. A comparison of Tables 8 and 9 indicates, MAP, PA, SCREE and K1 showed substantial improvement at all levels of Complexity when the saturation was .8. At this saturation level, PA slightly underestimated at Complexity level 6. MAP slightly overestimated at Complexity levels 2, 5 and 6. SCREE slightly overestimated at all levels of Complexity and level 6 resulted in its largest overestimation. The BART method overestimated slightly or not at all at levels 1, 2 and 3. K1 markedly overestimated at Complexity levels 4, 5 and

6 as did BART. Levels 4, 5 and 6 all contain unique variables.

Tables 10 and 11 parallel Tables 8 and 9 with P equal to 72. As was the case when P was 36, the range of complexity appears to have differentially affected the method's performance. At a saturation of .5, in Table 10, MAP was quite accurate at levels 1, 2, 4 and 5 but underestimated erratically at levels 3 and 6. At a saturation of .8, in Table 11, MAP was very accurate at all levels of complexity. PA was quite accurate across all levels of complexity with marked improvement at the .8 saturation level. At that level, PA was always accurate. Generally, The SCREE method somewhat overestimated at each level of complexity. It performed worst when the saturation was .5 at levels 2 and 4 and when the saturation was .8, level 4. The K1 method gave gross overestimates at all levels of complexity when the saturation was .5. It was quite accurate when the saturation was .8 at levels 1, 2 and 3. At the same saturation at levels 4, 5 and 6, the method consistently overestimated the criterion. The BART method showed a moderate range of underestimation when the saturation was .5 with the worst case appearing to be level 6. When the saturation was .8, BART performed well at levels 1 and 3, overestimated moderately at level 2 and overestimated greatly at levels 4, 5 and 6.

Insert Tables 10 and 11 About Here

A general overview of the performance of the different methods may be gained by calculating the percent of times each method's estimate deviated a set amount from the criterion. Since P and saturation appear to have had the most substantial impact on the methods, the percentages were computed at each level of these variables. Deviations of greater than three were collapsed for simplicity of presentation. Differences of 0 indicate accurate estimates.

These percentages are presented in Tables 12 and 13.

Insert Table 12 About Here

As Table 12 indicates, at both levels of saturation when P was 36, PA was clearly the most frequently accurate method followed by MAP and SCREE. K1's tendency to overestimate was marked. The K1 method never underestimated. The BART method was quite inaccurate and variable at both levels of saturation.

Table 13 parallels Table 12 with P equal to 72. Again, PA was the most frequently accurate method at both levels of saturation. Both the PA and the MAP methods showed improved performance when P was equal to 72 compared to 36. The PA and MAP methods were nearly equivalent when the saturation was .8. The K1 method showed essentially the same pattern of results when P was 72 as at 36 because of attenuated range on these tables. The BART method retained more components when P was 72 than 36. BART was more often accurate when the saturation was .5 than .8 when P was 72.

Insert Table 13 About Here

Discussion

The question of interest in this study was the ability of five decision methods to estimate the number of major components present in the population correlation matrices given only the generated sample matrices. The difference between the estimated number and the defined number of major components served as the primary dependent variable in this simulation study. The standard deviation of the difference scores gave further information about each

method's consistency. Finally, the percent of correct decisions and the percent of decisions at specified levels of deviation from the criterion were also calculated.

The five decision rules employed were the eigenvalue greater than one rule (K1), Bartlett's test (BART), the scree test (SCREE), the minimum average partial method (MAP) and the parallel analysis method (PA). The performance of the five methods for determining the number of components was examined in ten samples drawn from each of 48 simulated population correlation matrices over a range of component pattern complexity. The least complex pattern replicated earlier work (Zwick & Velicer, 1982) and included only equal numbers of variables per component and no unique or complex variables. Component pattern complexity was varied by modifying this clear, least complex case with combinations that included: (1) complex variables, (2) equal and unequal numbers of variables per component, and (3) unique variables. The resultant six levels of complexity are felt to cover an adequately wide range to permit a test of the relative strengths and weaknesses of the decision methods examined. Major components (MJC) were defined as those having more than three substantial loadings and an eigenvalue greater than or equal to 1.0 at the population level. Two types of minor components (MNC) were defined. It is felt that these complex patterns expand upon the formal model and incorporate cases likely to be encountered in real data analyses.

The K1 rule was found to overestimate consistently the number of major components. It never underestimated. This finding is consistent with those of Cattell and Jasper (1967), Linn (1968), Yeomans and Golden (1982) and Zwick and Velicer (1982). At a component saturation of .5, the number retained often fell in the 1/3 to 1/2 of P range discussed by Gorsuch (1974). As the number of variables increased, so did the number of components retained. K1

retained more components when unique variables were included in the population pattern. These findings are clearly contrary to those of Humphreys (1964) and Mote (1970), who concluded the K1 method sometimes retained too few components. Either their data represented a type of component complexity not included in this study or their original judgments of the number of components in their data sets were overestimates. Given the apparent functional relationship of the number of components retained by K1 to the number of variables and the repeated reports of the method's inaccuracy, the K1 rule cannot be recommended for PCA.

The results and conclusions about the K1 rule presented here are consistant with both previous empirical studies and the theoretical criticism of the method. However, our conclusions are in sharp contrast to many recent textbooks where K1 is either the preferred or only method discussed (Afifi & Clark, 1984; Everitt & Dunn, 1983; Johnson & Wichern, 1982; Marascuilo & Levin, 1983; Tabachnick & Fidell, 1983). For example, Marascuilo & Levin (1983) are typical when they first discuss it with a caution ("In most cases, Kaiser's rule is quite workable, but . . . ", p. 237) but, in a later summary, flatly recommend it ("It is one we recommend.", p. 260). The use of the K1 rule as the default value in some of the standard computer packages (BMDP, SPSS) represents an implicit endorsement of the procedure, particularly with naive users. This pattern of explicit endorsement by textbook writers and implicit endorsement by computer packages, contrasted with empirical findings that the procedure is very likely to provide a grossly wrong answer, seems to guarantee that a large number of incorrect findings will continue to be reported.

The BART method's performance was the most variable of those examined. In addition to variability, the method was sensitive to a number of influences.

Increases in N, P and SAT as well as the use of conservative alpha levels and the presence of unique variables all lead to the retention of more components. The first four of these influences may be seen as affecting the statistical power of the Bartlett test. In data sets where the P-m eigenvalues were in fact equal at the population level, Zwick and Velicer (1982) found the method to be moderately accurate. In the broader range of complexity examined here, the test tended to retain both types of minor components defined above. Although examination of different alpha levels led to fewer or greater numbers of components retained, the accuracy and consistency of the method did not appear to be markedly improved by adjusting alpha levels with sample size (see Table 4) as was suggested by Horn and Engstrom (1979). Other factors present in this study appear to have had a greater influence on the performance of BART, across alpha levels, than did sample size alone.

The Bartlett test is accurate in answering statistical questions concerning the equality of eigenvalues (Bartlett, 1950; 1951). Researchers inclined to examine minor components, particularly early in the course of exploratory analysis, may find the method helpful. However, the Bartlett test cannot be recommended as a general method of determining the number of major components to retain. The tendency of the method to retain both minor and trivial components might reflect the basic logic of the test. Only true error should be expected to meet the requirement of equal eigen values. However, most researchers would not find minor or trivial components to be consistent with their implicit definition of a factor or component that is worthy of retention. Therefore, the test may work correctly but may not be an appropriate test for most applications.

This study did not investigate the maximum lilelihood test which assumes

the factor analysis model. The maximum likelihood test is based on a logic identical to the Bartlett test. Empirical investigations have found the same pattern of results with the likelihood test as we have reported with the Bartlett test (Hakstian, Rogers, & Cattell, 1982; Richman, personal communication). Again, the problem may be with an inappropriate formulation of test, rather than the performance of the test.

The SCREE method had moderate overall reliability when the mean of two trained raters was used. The correlation of the mean of those raters' decisions with an expert judge indicated fair overall agreement. Reports of rater reliability on the SCREE have ranged from very good (Cattell and Jaspers, 1967) to quite poor (Crawford and Koopman, 1979). This range may reflect either the training or the task complexity across research projects. The raters in this study showed greater agreement at higher than at lower component saturation levels. They showed greater agreement when there were more rather than fewer variables. Perhaps more importantly, the interrater reliability of the SCREE procedure had a fairly wide range across levels of complexity. The moderate reliability of the SCREE method is very problematic for the applied researcher. Unreliability at this point in the analysis may well expose a study to otherwise avoidable experimenter bias. In any case, applied researchers should note that reliability questions always arise in any use of the SCREE method.

In general, the SCREE method was more accurate and less variable than either the KI or BART method. This method was more accurate and less variable at the higher level of component saturation. Larger sample sizes also improved its accuracy when P was 72 and SAT was .5. Sample size did not appreciably affect SCREE at other levels of P or SAT. This effect of larger sample size is consistent with those reported elsewhere (Cliff and Pennell,

1967; Linn, 1968; Zwick and Velicer, 1982). The accuracy of the SCREE method was not affected by an increase in the number of variables examined. An increase in the average number of variables per component did not affect the method's accuracy. In an earlier study (Zwick and Velicer, 1982), utilizing non-complex matrices, the SCREE method performed better than MAP when the major components contained 6 or less variables and the saturation was low. This trend can be observed again under more complex conditions.

The SCREE method tended to overestimate rather than to underestimate when it deviated from the criterion value. As reflected in Tables 12 and 13, the SCREE was accurate about 57% of the time. When the SCREE method was in error, 90 percent of the errors were overestimations. The SCREE method appeared to be most variable at the low level of component saturation or when unique or complex variables were present. Nevertheless, even given its variability and tendency toward overestimation, the SCREE method seldom led to the retention of more than two components over the criterion value. Many experienced investigators routinely examine 1, 2 or 3 components above and below the estimate given by their favorite decision method. This practice, coupled with good judgment concerning interpretability may often result in appropriate solutions when the SCREE method is employed. This optimistic assertion notwithstanding, the ever present question of rater reliability, the tendency to overestimate and the apparent increased variability in the common case of unique or complex variables all argue against the exclusive use of the SCREE method. Given these drawbacks and the availability of other clearly superior methods, SCREE can no longer be recommended as the method of choice for determining the number of components in PCA.

In general, the MAP method was more often accurate and less variable than the K1, BART or SCREE methods. It showed an overall tendency to underestimate

the criterion. The MAP method was most accurate at the higher level of component saturation or when the average number of variables per component was large. Its performance was not markedly influenced by sample size within the range examined in this study. The MAP method was quite accurate under many conditions and, when accurate, showed little variability. In cases representing both a low level of saturation and a low number of variables per component, the MAP method consistently underestimated the number of major components. This effect can be most clearly seen in Table 6 by comparing the two levels of P/MJC when P = 36 and SAT = .5. Additional information about this effect can be gleaned from Table 7. The MAP method gave larger underestimates and displayed greater variability when unique variables were present (levels 4, 5 and 6) and when there was an unequal number of variables per component (level 3 and 6). In this simulation study, unique variables reduced the number of variables per component by 1. The presence of unequal numbers of variables per component independently reduced the number of variables per component on the trailing major components. The presence of complex variables (level 2 and 5) lowers the major component saturation at the sample level. Complex variables thus independently increase the effect of low component saturation. The combination of these influences appears to result in components at the sample level which account for less variance than those components containing only a unique variable. MAP should not retain a unique variable component. It, therefore, fails to pick up the less well identified major components.

Overall, the MAP method was accurate more often than were the K1, BART or SCREE methods. The MAP method gave results within ± 1 of the criterion between 72% (P = 36, SAT = .5) and 100% (P = 36, SAT = .8; P = 72, SAT = .8) of the time. When it was in error, the MAP method tended to underestimate.

Approximately 90% of the MAP errors were underestimations.

The MAP method provides an unequivocal stopping point. It is tied to the concept of parsimony by directly rejecting components identified by only one variable. MAP showed a tendency to underestimate the known number of major components at the component satuation level of .5 when up to six variables loaded on a component. It is quite accurate when component saturation is high or when there is an average of 8 or more variables per component. Researchers wishing to ignore relatively small major components should employ MAP as a primary method of determining the number of components to retain.

The PA method was consistently accurate. It was typically the most accurate method at each level of complexity examined. The average deviation of PA from the criterion did not exceed exceed 0.30 under any condition examined. The difference scores of the PA method showed less variability than those of any other rule. Increases in sample size, component saturations and P/MJC improved the PA method's performance when there was room for improvement. It might have been expected that the PA method would overestimate in the presence of minor components made up of unique or complex variables. Some evidence of this is present at P = 72, SAT = .5, complexity levels 2, 4 and 5. This pattern is not replicated at other levels of SAT or P.

Overall, the PA method was the most frequently accurate method examined. It gave results within ± 1 of the criterion for between 95% (P = 36, SAT = .5) and 100% (P = 36, SAT = .8; P = 72, SAT = .8) of the cases examined. When the PA method was in error, it showed a slight tendency toward overestimation. Approximately 65% of the PA method's errors were overestimations. The accuracy of the PA method in this study is consistent with the CFA results reported by Humphreys and Montanelli (1975).

A major drawback in the applied use of the PA method is the necessity of

generating a large set of random correlation matrices at the particular combination of P and N under consideration. The prediction equation developed for principal axis factor analysis (Montinelli and Humphreys, 1976) is not appropriate for principal components analysis.

In summarizing the results of the present study it is useful to postulate a further division of the class of component previously labeled as Major Component (MJC). Those components which involve only a limited number of variables and low saturation will be labelled as Poorly Defined Components (PDC). Components with either a large number of variables or high saturation will be labelled as Well Defined Components (WDF). Poorly Defined Components do not possess any "marker" varibles, i.e., variables with high loadings on that component. Investigators typically employ such marker variables as defining varibles in interpreting the component. Guadagnoli and Velicer (1984) found that PDC's were unlikely to accurately replicate even in fairly large samples. The combination of the two issues of difficult interpretability and questionable replicability, make the retention of these components problematic.

The two methods which were the most accurate, MAP and PA, provided divergent results primarily when PDC's were present. An A Priori decision about whether or not to attempt to extract and retain such components may dictate whether MAP or PA is the method of choice. Lacking such a decision, a researcher may want to examine all solutions in a set bracketted by the MAP and PA estimates. Rotational criteria and interpretability may be the basis for a final decision.

Previous studies have examined subsets of these rules under some of the conditions examined here. The present study provides comparisons across a wider variety of situations than previous investigations. In those areas

where the simulated situations were similar, the results of Linn (1968), Humphreys and Montanelli (1975), Cattell and Vogelman (1977), Hakstian, Rogers, and Cattell (1982) and Zwick and Velicer (1982) were confirmed and expanded.

Within the limitations imposed by the simulation approach, the results of this study, paired with previously reported work, permit some conclusions concerning methods of determining the number of components in real data sets. There is no evidence supporting the continued use of K1 or BART as exclusive, primary methods to determine the number of major components to retain. These methods should no longer be employed. The SCREE procedure has been reported to be relatively accurate. This study is consistent with those reports but indicates that the method is too variable and too likely to overestimate to be employed as the sole decision method. However, the SCREE may still be useful for initial estimates or as a complementary method employed in conjunction with PA or MAP. The MAP method was generally quite accurate and consistent when the component saturation was high or the component was defined by more than six variables. The MAP did not retain Poorly Defined Components. The PA method was also consistently accurate. PA retained PDC's and showed a slight tendency to overestimate. The general application of the PA method is difficult to recommend because programs needed for its application are not widely available. Assuming that these problems can be overcome, either PA or MAP are the method of choice, with many situations arising where both should be employed.

Footnote

Some of the material included in the present paper appeared in the senior author's Ph.D. Dissertation. Reprint requests should be sent to William R. Zwick, Counseling Department, Rhode Island Group Health Association, 530 N. Main Street, Providence, RI 02906.

References

Afifi, A.A., & Clark, V. (1984). Computer-aided multivariate analysis. Belmont, CA: Lifetime Learning Publications.

Anderson, A. D., Acito, F. & Lee, H. (1982). A simulation study of three methods for determining the number of image components. Multivariate Behavioral Research, 17, 493-502.

Anderson, T. W. & Rubin, H. (1956). Statistical inference in factor analysis. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 5.

Bartlett, M. S. (1950). Tests of significance in factor analysis. The British Journal of Psychology, 3, 77-85.

Bartlett, M. S. (1951). A further note on tests of significance in factor analysis. The British Journal of Psychology, 4, 1-2.

Browne, M. W. (1968). A note on lower bounds for the number of common factors. Psychometrika, 33, 233.

Cattell, R. B. (1966). The scree test for the number of factors. Multivariate Behavioral Research, 1, 245-276.

Cliff, N. (1970). The relation between sample size and population characteristic Vectors. Psychometrika, 35, 163-178.

Cattell, R. B. & Jaspers, J. (1967). A general plasmode for factor analytic exercises and research. Multivariate Behavioral Research Monographs, 3, 1-212.

Cattell, R. B. & Vogelman, S. (1977). A comprehensive trial of the scree and KG criteria for determining the number of factors. Multivariate Behavioral Research, 12, 289-325.

Chatfield, C., & Collins, A.J. (1980). Introduction to multivariate analysis. London: Chapman and Hall.

Cliff, N. & Hamberger, C. D. (1967). The study of sampling errors in facator analysis by means of artificial experiments. Psychological Bulletin, 68, 430-445.

Cliff, N. & Pennell, R. (1967). The influence of communality, factor strength and loading size on the sample characteristics of factor loadings. Psychometrika, 32,, 309.

Comrey, A.L. (1978). Common methodological probelms in factor analytic studies. Journal of Consulting and Clinical Psychology, 46, 648-659.

Crawford, C. B. (1975). Determining the number of interpretable factors. Psychological Bulletin, 82, 2, 226-237.

Crawford, C. B. & Koopman, P. (1979). Note: Inter-rater reliability of scree test and mean square ratio test of number of factors. Perceptual and Motor Skills, 49, 223-226.

Everett, J. E. (1983). Factor comparability as a means of determining the number of factors and their rotation. Multivariate Behavioral Research, 18, 197-218.

Everitt, B.S., & Dunn, G. (1983). Advanced methods of data exploration and modeling. London: Heinemann.

Glass, G. V. & Taylor, P. A. (1966). Factor analytic methodology. Review of Educational Research, 36, 566-587.

Gorsuch, R. L. (1974). Factor Analysis. Philadelphia: Saunders.

Gorsuch, R.L. (1973). Using Bartlett's significance test to determine the number of factors to extract. Educational and Psychological Measurement, 33, 361-364.

Green, S. B. (1983). Identifiability of spurious factors using linear factor analysis with binary items. Applied Psychological Measurement, 7, 139-147.

Guadagnoli, E., & Velicer, W.F. (1984). Stability of component patterns: A simulation study. Unpublished manuscript.

Guttman, L. (1954). Some necessary conditions for common factor analysis. Psychometrika, 19, 149-162.

Guttman, L. (1955). the determinancy of factor score matrices with implications for five basic problems of common-factor theory. British Journal of Statistical Psychology, 8, 65 - 81.

Hakstian, A. R., Rogers, W. T. & Cattell, R. B. (1982). The behavior of number of factors rules with simulated data. Multivariate Behavioral Research, 17, 193-219.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. Psychometrika, 30, 179-185.

Horn, J. L. (1969). On the internal consistency reliability of factors. Multivariate Behavioral Research, 4, 115-125.

Horn, J. L. & Engstrom, R. (1979). Cattell's scree test in relation to Bartlett's chi-square test and other observations on the number of factors problem. Multivariate Behavioral Research, 14, 283-300.

Horst, P. (1965). Factor Analysis of Data Matrices. New York: Holt, Rinehart and Winston.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24, 417-441, 498-520.

Humphreys, L. G. (1964). Number of cases and number of factors: An example where N is very large. Educational and Psychological Measurement, 24, 457.

Humphreys, L. G. & Montanelli, R. G. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. Multivariate Behavioral Research, 10, 193-206.

Jackson, B.B. (1983). Multivariate data analysis. Homewood, IL: Richard D. Irwin.

Jackson, D. N. & Chan, D. W. (1980). Maximum-Likelihood estimation in common factor analysis: A cautionary note. Psychological Bulletin, 88, 502-508.

Johnson, R.A., & Wichern, D.W. (1982). Applied multivariate statistical analysis. Englewood Cliffs, NJ: Prentice-Hall.

Joreskog, K.G. (1962). On the statistical treatment of residuals in factor analysis. Psychometrika, 27, 335-354.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20, 141-151.

Kaiser, H. F. (1970). A second-generation little jiffy. Psychometrika, 35, 401-415.

Kaiser, H. F. & Caffrey, J. (1965). Alpha factor analysis. Psychometrika, 30, 1-44.

Lawley, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. Proceedings of the Royal Society of Edinburgh, 60, 64-82.

Lawley, D.M. (1941). further investigation in factor estimation. Proceedings of the Royal Society of Edinburgh, 61, 176-185.

Lee, H. B. & Comrey, A. L. (1979). Distortions in a commonly used factor analytic procedure. Multivariate Behavioral Research, 14, 301-321.

Linn, R. L. (1968). A Monte Carlo approach to the number of factors problem. Psychometrika, 33, 37-71.

Humphreys, L. G. & Montanelli, R. G. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. Multivariate Behavioral Research, 10, 193-206.

Jackson, B.B. (1983). Multivariate data analysis. Homewood, IL: Richard D. Irwin.

Jackson, D. N. & Chan, D. W. (1980). Maximum-Likelihood estimation in common factor analysis: A cautionary note. Psychological Bulletin, 88, 502-508.

Johnson, R.A., & Wichern, D.W. (1982). Applied multivariate statistical analysis. Englewood Cliffs, NJ: Prentice-Hall.

Joreskog, K.G. (1962). On the statistical treatment of residuals in factor analysis. Psychometrika, 27, 335-354.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20, 141-151.

Kaiser, H. F. (1970). A second-generation little jiffy. Psychometrika, 35, 401-415.

Kaiser, H. F. & Caffrey, J. (1965). Alpha factor analysis. Psychometrika, 30, 1-44.

Lawley, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. Proceedings of the Royal Society of Edinburgh, 60, 64-82.

Lawley, D.M. (1941). further investigation in factor estimation. Proceedings of the Royal Society of Edinburgh, 61, 176-185.

Lee, H. B. & Comrey, A. L. (1979). Distortions in a commonly used factor analytic procedure. Multivariate Behavioral Research, 14, 301-321.

Linn, R. L. (1968). A Monte Carlo approach to the number of factors problem. Psychometrika, 33, 37-71.

Lunneborg, C.E., & Abbott, R.D. (1983). Elementary multivariate analysis for the behavioral sciences. New York: North-Holland.

Marasculiolo, L.A., & Levin, J.R. (1983). Multivariate statistics in the social sciences: A researcher's guide. Monterey, CA: Brooks/Cole.

Montanelli, R. G., Jr. (1975). A computer program to generate sample correlation and covariance matrices. Educational and Psychological Measurement, 35, 195-197.

Montanelli, R. G. & Humphreys, L. G. (1976). Latent roots of random data correlation matrices with squared multiple correlations on the diagonal: A Monte Carlo study. Psychometrika, 41, 3, 341-348.

Morrison, D.F. (1976). Multivariate Statistical Methods (Second Ed.) New York: McGraw-Hill.

Mote, T. A. (1970). An artifact of the rotation of too few factors: Study orientation vs. trait anxiety. Revista Interamericana de Psicologia, 37, 61-91.

Pruzek, R.M. & Rabinowitz, S.N. (1981). A simple method for exploratory structural analysis. American Educational Research Journal, 18, 173-189.

Revelle, W. & Rocklin, T. (1979). Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. Multivariate Behavioral Research, 14, 403-414.

Schonemann, P. H. & Steiger, J. H. (1978). On the validity of indeterminant factor scores. Bulletin of the Psychonomic Society, 12, 287-290.

Schonemann, P. H. & Wang, M. (1972). Some new results on factor indeterminacy. Psychometrika, 37, 61-91.

Steiger, J. H. & Schonemann, P. H. (1979). A history of factor indeterminacy. In Shye, S. (Ed.) Theory Construction and Data Analysis in the Behavioral Sciences. San Francisco: Jossey-Bass.

Tabachnick, B.G., & Fidell, L.S. (1983). Using multivariate statistics. New York: Harper & Row.

Tucker, L. B., Koopman, R. F. & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. Psychometrika, 34, 421.

Van de Geer, J. P. (1971). Introduction to Multivariate Analysis for the Social Sciences. San Francisco: Freeman.

Veldman, D. J. (1979). Simple structure and the number of factors problem. Multivariate Behavioral Research, 14, 191-200.

Velicer, W. F. (1974). A comparison of the stability of factor analysis, principal component analysis and rescaled image analysis. Educational and Psychological Measurement, 34, 563-572.

Velicer, W. F. (1976a). Determining the number of components from the matrix of partial correlations. Psychometrika, 41, 3, 321-327.

Velicer, W. F. (1976b). The relation between factor score estimates, image scores and principal component scores. Educational and Psychological Measurement, 36, 149-159.

Velicer, W. F. (1977). An empirical comparison of the similarity of principal component, image, and factor patterns. Multivariate Behavioral Research, 12, 3-22.

Velicer, W.F., & Jackson, D.N. (1984). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. Unpublished manuscript.

Velicer, W. F., Peacock, A. C. & Jackson, D. N. (1982). A comparison of component and factor patterns: A Monte Carlo approach. Multivariate Behavioral Research, 17, 371-388.

Yeomans, K. A. & Golder, P. A. (1982). The Guttman-Kaiser criterion as a predictor of the number of common factors. The Statistician, 31, 221-229.

Zwick, W. R. & Velicer, W. F. (1982). Factors influencing four rules for determining the number of components to retain. Multivariate Behavioral Research, 17, 253-269.

Table 1

The Relationship Between the Number of Substantial Component Loadings and Eigenvalue Size to Major, Minor and Trivial Components.

Variables Per Component	Eigenvalue	
	Less Than 1.0	Greater Than 1.0
Less Than 3	Trivial (TC)	Minor (MNC)
Greater Than 3	Minor (MNC)	Major (MJC)

Table 2

Number of Variables Loading Substantially on Each Component under the Condition of Unequal Variables/Component for Different Values of P and Numbers of Major Components (MJC).

			Component								
P	MJC	P/MJC	1	2	3	4	5	6	7	8	9
36	3	12	14	12	10	-	-	-	-	-	-
36	6	6	8	7	6	6	5	4	-	-	-
72	6	12	15	14	13	11	10	9	-	-	-
72	9	8	12	11	10	9	8	7	6	5	4

Table 3

Interrater reliability of the Trained Scree Raters with Each Other at Each Level of the Number of Variables, Pattern Complexity and Component Saturation, Correct for the Number of Raters.

	Saturation	
Complexity	.50	.80
	P = 36 Variables	
1	.67	.96
2	.96	.95
3	.82	.98
4	.76	.97
5	.65	.91
6	.77	.77
	P = 72 Variables	
1	.97	.99
2	.80	.97
3	.95	1.00
4	.93	.81
5	.78	.82
6	.75	.61

Table 4

Means and Standard Deviations of the Difference from the Population MJC at Different Sample Sizes when P = 36.

Sample Size		Method						
		MAP	PA	SCREE	K1	BA	BB	BC
		Saturation = .5						
72	d	-1.08	-0.05	0.50	8.32	-2.87	-3.92	-3.98
	(sd)	(1.68)	(0.70)	(1.04)	(1.21)	(1.10)	(0.81)	(0.79)
180	d	-1.17	0.13	0.68	7.30	-1.78	-2.20	-2.27
	(sd)	(1.84)	(0.39)	(0.95)	(1.09)	(1.26)	(0.97)	(0.92)
		Saturation = .8						
72	d	0.10	-0.02	0.27	1.77	0.47	-0.48	-0.60
	(sd)	(0.30)	(0.13)	(0.50)	(1.59)	(2.81)	(2.31)	(2.30)
180	d	0.0	0.0	0.23	1.32	1.23	0.68	0.62
	(sd)	(0.0)	(0.0)	(0.52)	(1.44)	(3.51)	(3.27)	(3.23)

Table 5

Means and Standard Deviations of the Difference from the Population MJC at Different Sample Sizes when P = 72.

Sample Size		Method						
		MAP	PA	SCREE	K1	BA	BB	BC
		Saturation = .5						
144	d	-0.45	0.02	1.16	17.80	-0.43	-1.60	-1.73
	(sd)	(1.00)	(0.57)	(1.30)	(1.29)	(1.16)	(1.24)	(1.15)
360	d	-0.43	0.07	0.46	15.42	0.40	-0.13	-0.22
	(sd)	(1.06)	(0.25)	(0.79)	(1.27)	(0.74)	(0.43)	(0.45)
		Saturation = .8						
144	d	0.02	0.0	0.28	2.97	3.88	2.62	2.50
	(sd)	(0.13)	(0.0)	(0.55)	(2.81)	(3.64)	(2.96)	(2.88)
360	d	0.0	0.0	0.31	2.52	5.03	4.10	3.98
	(sd)	(0.0)	(0.0)	(1.03)	(2.59)	(4.31)	(3.89)	(3.78)

Table 6

Means and Standard Deviations of the Difference from the Population MJC at Different Numbers of Variables Per Component when P = 36.

P/MJC		Method						
		MAP	PA	SCREE	K1	BA	BB	BC
		Saturation = .5						
6	d	-2.27	0.05	0.65	7.10	-1.50	-2.67	-2.77
	sd	(1.88)	(0.79)	(1.19)	(0.99)	(1.35)	(1.58)	(1.54)
12	d	0.02	0.03	0.53	8.52	-3.15	-3.45	-3.48
	(sd)	(0.13)	(0.18)	(0.75)	(1.10)	(0.44)	(0.53)	(0.57)
		Saturation = .8						
6	d	0.08	0.17	0.33	1.68	2.93	2.17	2.10
	(sd)	(0.28)	(0.13)	(0.59)	(1.70)	(2.94)	(2.54)	(2.54)
12	d	0.02	0.0	0.17	1.40	-1.23	-1.97	-2.08
	(sd)	(0.13)	(0.0)	(0.39)	(1.33)	(1.73)	(1.24)	(1.25)

Table 7

Means and Standard Deviations of the Difference from the Population MJC at Different Numbers of Variables Per Component when P = 72.

P/MJC		Method						
		MAP	PA	SCREE	K1	BA	BB	BC
		Saturation = .5						
8	d	-0.92	0.07	1.02	15.90	-0.17	-1.37	-1.48
	(sd)	(1.28)	(0.61)	(1.22)	(1.45)	(1.21)	(1.36)	(1.32)
12	d	0.03	0.02	0.60	17.32	0.13	0.37	0.47
	(sd)	(0.18)	(0.13)	(1.00)	(1.74)	(0.85)	(0.66)	(0.65)
		Saturation = .8						
8	d	0.02	0.0	0.21	3.02	5.15	4.00	3.87
	(sd)	(0.13)	(0.0)	(0.74)	(3.02)	(4.56)	(4.04)	(3.93)
12	d	0.0	0.0	0.38	2.47	3.77	2.72	2.62
	(sd)	(0.0)	(0.0)	(0.89)	(2.34)	(3.27)	(2.80)	(2.73)

Table 8

Means and Standard Deviations of the Difference from the Population MJC at Different Levels of Pattern Complexity when P = 36 and Saturation = .5.

Complexity		Method						
		MAP	PA	SCREE	K1	BA	BB	BC
1	d	-0.30	0.0	0.38	7.15	-0.90	-1.15	-1.20
	(sd)	(0.66)	(0.32)	(0.92)	(1.46)	(1.29)	(1.46)	(1.51)
2	d	-0.50	-0.10	0.88	7.35	-1.00	-1.40	-1.45
	(sd)	(0.89)	(0.45)	(0.93)	(1.22)	(1.34)	(1.54)	(1.54)
3	d	-0.80	-0.10	0.50	7.30	-1.20	-1.40	-1.55
	(sd)	(1.06)	(0.45)	(1.03)	(1.03)	(1.24)	(1.27)	(1.39)
4	d	-1.60	0.0	0.38	8.45	-1.55	-1.85	-1.90
	(sd)	(2.09)	(0.65)	(0.93)	(1.10)	(1.79)	(1.84)	(1.80)
5	d	-2.05	-0.25	0.58	8.20	-1.30	-1.85	-1.90
	(sd)	(2.50)	(0.79)	(0.96)	(0.95)	(1.59)	(1.93)	(1.89)
6	d	-1.50	0.20	0.85	8.40	-1.30	-1.70	-1.75
	sd	1.96	0.62	1.18	1.10	1.42	1.66	1.68

Table 9

Means and Standard Deviations of the Difference from the Population MJC at Different Levels of Pattern Complexity when P = 36 and Saturation = .8.

Complexity		Method						
		MAP	PA	SCREE	K1	BA	BB	BC
1	d	0.0	0.0	0.12	0.15	0.05	0.0	0.0
	(sd)	(0.0)	(0.0)	(0.32)	(0.37)	(0.22)	(0.0)	(0.0)
2	d	0.20	0.0	0.25	0.15	0.30	0.10	0.0
	(sd)	(0.41)	(0.0)	(0.52)	(0.37)	(0.57)	(0.31)	(0.0)
3	d	0.0	0.0	0.10	0.10	0.05	0.0	0.0
	(sd)	(0.0)	(0.0)	(0.26)	(0.31)	(0.22)	(0.0)	(0.0)
4	d	0.0	0.0	0.32	2.90	3.20	2.70	2.50
	(sd)	(0.0)	(0.0)	(0.69)	(0.85)	(1.61)	(1.72)	(1.88)
5	d	0.05	0.0	0.32	2.85	4.55	3.80	3.65
	(sd)	(0.22)	(0.0)	(0.57)	(0.74)	(2.06)	(2.02)	(1.81)
6	d	0.05	-0.05	0.38	3.10	3.45	3.00	2.90
	(sd)	(0.22)	(0.22)	(0.54)	(0.64)	(1.85)	(1.92)	(1.97)

Table 10

Means and Standard Deviations of the Difference from the Population MJC at Different Levels of Pattern Complexity when P = 72 and Saturation = .5.

Complexity		Method						
		MAP	PA	SCREE	K1	BA	BB	BC
1	d	0.0	0.0	0.82	15.95	-0.25	-0.60	-0.75
	(sd)	(0.46)	(0.0)	(1.09)	(1.82)	(0.55)	(0.82)	(0.79)
2	d	-0.05	0.05	1.18	16.00	-0.15	-0.60	-0.65
	(sd)	(0.22)	(0.22)	(1.08)	(2.10)	(0.37)	(0.68)	(0.67)
3	d	-0.95	-0.20	0.40	15.90	-0.60	-1.00	-1.10
	(sd)	(1.19)	(0.41)	(0.75)	(1.71)	(1.31)	(1.38)	(1.25)
4	d	-0.10	0.30	1.00	17.15	-0.45	-0.95	-1.10
	(sd)	(0.45)	(0.57)	(1.48)	(1.46)	(0.94)	(1.19)	(1.25)
5	d	-0.10	0.20	0.90	17.15	-0.40	-0.85	-1.00
	(sd)	(0.45)	(0.41)	(1.11)	(1.39)	(9.99)	(1.22)	(1.21)
6	d	-1.45	-0.10	0.35	17.50	-0.70	-1.20	-1.25
	(sd)	(1.60)	(0.55)	(0.97)	(1.24)	(1.45)	(1.58)	(1.55)

Table 11

Means and Standard Deviations of the Difference from the Population MJC at Different Levels of Pattern Complexity when P = 72 and Saturation = .8.

Complexity		Method						
		MAP	PA	SCREE	K1	BA	BB	BC
1	d	0.0	0.0	0.12	0.0	0.05	0.0	0.0
	(sd)	(0.0)	(0.0)	(0.32)	(0.0)	(0.22)	(0.0)	(0.0)
2	d	0.05	0.0	0.18	0.35	1.70	0.90	0.80
	(sd)	(0.22)	(0.0)	(0.41)	(0.59)	(1.22)	(0.91)	(0.89)
3	d	0.0	0.0	0.0	0.05	0.05	0.0	0.0
	(sd)	(0.0)	(0.0)	(0.0)	(0.22)	(0.22)	(0.0)	(0.0)
4	d	0.0	0.0	1.05	5.40	5.85	5.30	5.20
	(sd)	(0.0)	(0.0)	(1.31)	(0.94)	(2.03)	(2.05)	(1.96)
5	d	0.0	0.0	0.40	5.40	7.95	7.10	6.80
	(sd)	(0.0)	(0.0)	(0.50)	(0.82)	(2.16)	(2.07)	(1.99
6	d	0.0	0.0	0.0	5.25	7.60	6.85	6.65
	(sd)	(0.0)	(0.0)	(1.08)	(0.85)	(2.76)	(2.56)	(2.64)

Table 12

Percent of Each Method's Estimate Deviating a Set Amount from the Population MJC when P = 36.

Deviation	Method						
	MAP	PA	SCREE	K1	BA	BB	BC
	Saturation = .5						
±3	0.0	0.0	4.1	100.0	0.0	0.0	0.0
+2	0.0	1.7	16.7	0.0	0.0	0.0	0.0
+1	0.8	10.0	33.3	0.0	1.7	0.0	0.0
±0	59.2	82.5	36.7	0.0	9.2	2.5	1.7
-1	11.7	2.5	6.7	0.0	19.2	11.7	10.8
-2	9.2	3.3	2.5	0.0	10.0	12.5	12.5
±-3	19.2	0.0	0.0	0.0	59.9	73.3	75.0
	Saturation = .8						
±3	0.0	0.0	0.8	36.7	25.8	17.5	17.5
+2	0.0	0.0	5.0	12.5	1.7	7.5	7.5
+1	5.0	0.0	20.8	7.5	14.2	0.8	0.8
±0	95.0	99.2	73.3	43.3	26.7	34.2	33.3
-1	0.0	0.8	0.0	0.0	6.7	6.7	6.7
-2	0.0	0.0	0.0	0.0	6.7	7.5	4.2
±-3	0.0	0.0	0.0	0.0	18.3	25.8	30.0

Table 13

Percent of Each Method's Estimate Deviating a Set Amount from the Population MJC when P = 72.

Deviation	Method						
	MAP	PA	SCREE	K1	BA	BB	BC
	Saturation = .5						
±3	0.0	0.0	9.2	100.0	1.7	0.0	0.0
+2	0.0	0.8	19.1	0.0	5.8	0.8	0.0
+1	2.5	8.3	21.7	0.0	10.8	1.7	1.7
±0	75.8	85.8	46.7	0.0	58.3	47.5	41.7
-1	6.7	4.2	2.5	0.0	16.7	25.8	31.7
-2	6.7	0.0	0.8	0.0	5.8	11.4	10.8
±-3	8.3	0.0	0.0	0.0	0.8	12.5	14.1
	Saturation = .8						
±3	0.0	0.0	5.0	50.0	60.0	47.5	45.8
+2	0.0	0.0	1.7	0.8	3.3	8.3	9.2
+1	0.8	0.0	22.5	5.0	9.2	3.3	3.3
±0	99.2	100.0	69.2	44.2	27.5	40.8	41.7
-1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-2	0.0	0.0	0.8	0.0	0.0	0.0	0.0
±-3	0.0	0.0	0.8	0.0	0.0	0.0	0.0